JUN 3 1981
Item 491-B-13
HE 20.6512/2:St8

NCHSR

RESEARCH SUMMARY SERIES

COMPLETED
ORIGINAL

Comparison of Strategies for the Development of Process Measures in Emergency Medical Services

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health, Research, Statistics, and Technology
National Center for Health Services Research

National Center for Health Services Research
Research Summary Series

The *Research Summary Series* is published by the National Center for Health Services Research (NCHSR) to provide rapid access to significant results of NCHSR-supported research projects. The series presents executive summaries prepared by the investigators. Specific findings are highlighted in a more concise form than in the final report. The *Research Summary Series* is intended for health services administrators, planners, and other research users who require recent findings relevant to immediate problems in health services.

Abstract

A comparison was made of the use of two group process techniques, Delphi and Nominal Group Technique, to achieve consensus from panels of Emergency Physicians concerning the steps comprising appropriate care of patients described in brief scenarios. The results suggest that of the two methods, Delphi produces more frequent, greater and more stable consensus, but the process is less satisfying to the physicians involved. The latter attribute was considered a source of delays by physicians in completing the Delphi responses.

NCHSR

RESEARCH SUMMARY SERIES

Comparison of Strategies for the Development of Process Measures in Emergency Medical Services

Carolyn Atkinson Thornell
Boone, Young and Associates, Inc.

March 1981

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Services Research

DHHS Publication No. (PHS) 81-3277

This *NCHSR Research Summary* was prepared by Carolyn Atkinson Thornell, Principal Investigator, Boone, Young and Associates, Inc., Management Consultants, New York, N.Y. The study was supported by NCHSR under contract HRA 230-75-0176.

The full report of the study may be purchased under the same title, from the National Technical Information Service, Springfield, VA 22161 (703/487-4650), and may be ordered in either paper or microfiche using the following order numbers: PB 81 170326, Final Report; PB 81 170334, Appendices and Attachments.

Additional copies of the *NCHSR Research Summary* may be obtained on request from the NCHSR Publications and Information Branch, 3700 East-West Highway, Room 7-44, Hyattsville, MD 20782 (301/436-8970). Current NCHSR publications are listed in the back of this publication.

Office of Health Research, Statistics, and Technology
Ruth S. Hanft, Deputy Assistant Secretary for
Health Research, Statistics, and Technology

National Center for Health Services Research
Gerald Rosenthal, Ph.D., Director
Barbara P. McCool, Ph.D., Deputy Director
Robert A. Fordham, Associate Deputy Director
Donald E. Goldstone, M.D., Associate Deputy Director,
Medical and Scientific Affairs

Division of Extramural Research
Archer C. Copley, Director (Acting)
William B. Lohr, Chief (Acting), Institutional Care/
Emergency Medical Services/Long-Term Care Cluster
Elinor Walker, Project Officer

The views expressed in this publication are those of the author, and no official endorsement by the National Center for Health Services Research is intended or should be inferred.

Foreword

This is a study in standard-setting in Emergency Medicine. At present, it appears that quality assessment, quality assurance, and the measurement of effectiveness of health services depend upon the ability of providers to agree upon standards for the care process: that is, what services are given, to whom, and how. The key to explicating these standards, whether they are used for purposes of research or program-evaluation, or whether they actually guide the delivery of care, is the consensus judgment of experts. In developing standards, it is useful to have standardized replicable, and defensible methods of reaching these judgments. Two such methods are Delphi and Nominal Group Process (NGP). These two methods, based upon rather different theories of decision-making, are compared in this study.

The comparisons of Delphi and NGP made in this study do not address which standards are "best"; a very different sort of study would be needed to address that question. They do examine the utility of each technique in achieving consensus, the stability of the consensus, and certain related considerations having to do with how physicians accept their roles as participants.

With regard to the latter issue, a problem arises in interpreting the findings. Delphi was found to produce more frequent and more stable consensus than NGP. It was also less appealing to the participants: in Delphi the physicians remain anonymous; in NGP they interact face-to-face. They expressed their relative dissatisfaction directly in their comments, but they also tended to delay their responses. One consequence of this delay is a difference of about one year in the time at which the NGP results were achieved and the final Delphi results. This delay raises the possibility that "maturation," defined as a threat to the validity of the research results "as a function of the passage of time per se," [1] might plausibly have produced changes in the ability of physi-

[1] Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally College Publishing Company, 1963.

cians to reach agreement, the state of medical practice, or some other factor independent of the type of procedure used.

Those who are contemplating the use of a group-judgment technique for standard-setting are invited to examine the methods used here and the likelihood that maturation has affected the results. In view of the problem of Delphi by mail, as it was carried out here, they may wish to explore the feasibility of using Delphi in a computer-aided system, whereby the Delphi rounds would be conducted sequentially in an abbreviated time sequence. The physician-participants could be brought to a central location, though they would still respond in a noninteractive mode. This or some other method of reducing dependency upon mailed responses and repeated follow-up might offer the benefits of Delphi without the attendant delay, and might even bring the cost of the procedure to a minimum.

Additional detail on this study can be obtained from the National Technical Information Service (PB 81 170326, and PB 81 170334).

Gerald Rosenthal, Ph.D
Director
March 1981

Contents

1 Introduction

3 Study design and methodology

7 Findings
Qualitative analysis of the Delphi and NGT group processes
9 Number of procedures generated with each technique
10 Level of consensus achieved
11 Stability of Delphi and NGT votes over rounds
13 Commonality of consensus items across panels
16 Comparative costs associated with implementing Delphi and NGT

Conclusion

BLANK PAGE

BLANK PAGE

Introduction

In 1975, the National Center for Health Services Research (NCHSR) of what was then the Department of Health, Education, and Welfare contracted with Boone, Young and Associates, Inc. to Compare Strategies for the Development of Process Measures in Emergency Medical Services (EMS). Under this contract, the task was to design and implement under controlled conditions a study in which selected panels of physician-experts would be designated to use either the Delphi Technique (a non-interactive group judgment method for reaching consensus in group decision making) or the Nominal Group Technique (an interactive group judgment method) to reach consensus on steps in the process of care which should be followed in the handling of specific emergency medical cases. The nature and extent of any differences in the recommendations developed by these two techniques were to be analyzed to determine whether such results were attributable to the particular technique used and whether one method was more useful than the other for generating consensus with respect to procedures for handling emergency cases. The quality of the content of the process measures themselves was not to be evaluated. Rather the study was to focus on differences between the two techniques in the consensus reached on the process measures.

The Delphi and Nominal Group Techniques were the methods specified by NCHSR for evaluation and comparison in this study because of the success with which each had been used for decision making in a number of settings, including health-related areas. However, at the time this study was undertaken, the only study known to have used either Delphi or NGT for the development of process steps in EMS was the one Boone, Young and Associates had done in 1974.

In that study, physicians and nurses who were experts in their specialty and/or in the delivery of emergency medical care used the Delphi Technique to develop consensus as to the core process

steps which should be followed in handling ten emergency medical cases. The findings of that study indicated that the consensus reached might have been different and/or more stable if the cases about which recommendations were generated had been better delineated; that the consensus reached might have been different and/or more common across panels if all participants had been emergency medical specialists; and that the consensus reached through Delphi did not always hold up in face-to-face discussion of the cases by the same panel participants. In short, the findings suggested that, if the identification of an appropriate methodology for arriving at a consensus about criteria for adequate emergency care was the goal, the usefulness of the Delphi Technique for that purpose should be assessed under more controlled circumstances, and further that Delphi's value should be experimentally compared to that of a group judgment technique in which the experts interacted in the course of attempting to reach a consensus.

These findings suggestive of further study coincided with NCHSR's recognition, articulated in the Scope of Work for this contract, that "the development of PSROs makes imminent the widespread promulgation of 'expert-judgment' derived, process-based, measurements of quality." The Request for Proposal went on to say:

> At the moment, there is need for such measures and no adequate substitute for expert judgments, based upon clinical experience and training, as their source. Nevertheless, generation of such judgments through methodologies that remain unexamined and unsystematic is inconsistent with the development of data for quality assurance. It is necessary to explore and document efficient methods of developing reliable and acceptable guidelines for carrying out the medical care process. To do so, it will be necessary to compare the end products of a number of different decision making modes and structures in order to identify patterns that produce reliable results.

This project, therefore, represents part of the effort being undertaken by NCHSR to identify and compare the effectiveness of strategies for decision making in the area of emergency medical care.

Study design and methodology

Given the evaluative character of the study, a study design was required such that the effectiveness of the experimental variable—the group judgment mode—in generating consensus about medical care could be systematically assessed. Such a study design should allow for the rigorous control of all significant variables except the experimental variable, thus permitting any difference in outcome to be attributed to the operation of the group decision technique.

The major variables in the study, apart from the group judgment technique, were thought to be the cases to be dealt with and the composition of the panels of experts. Several design requirements were apparent with respect to controlling these. First, the cases for which the process measures were to be developed should be the same for both group decision situations. Second, the experts reviewing the cases should be drawn from the same universe. Third, the experts in one group decision situation should be matched as closely as possible to those in the other. Fourth, the case statements should be as unambiguous as possible so that variations in outcome could not be attributable to differences in case interpretation. Finally, several cases should be presented for discussion to two or more panels in each group decision situation so that the extent to which outcome was affected by the case under discussion and the particular panel making the decision could be assessed.

The basic research design therefore specified that several clearly delineated emergency medical cases would be presented to two or more Delphi panels and two or more NGT panels of physician-experts, drawn from the same universe and matched with respect to salient characteristics. Each panel would follow its assigned group decision technique to reach consensus about quality emergency management of each of the cases.

The sample was selected from a universe of emergency medical specialists, that is, physicians engaged primarily in the delivery of emergency medical services, or in the management of Emergency Departments. Panelists were nominated by other EMS specialists, and assigned randomly as either potential Delphi participants or potential NGT panelists. Thus inability to participate in one

technique precluded the possibility of participating in the other. From among the potential panelists invited and willing to participate, six panels—three Delphi and three NGT—were constituted which were matched as well as possible in terms of age, specialty, and geographical location, variables thought to be critical to control in this study. The three Delphi panels had nine members each; two NGT panels had six members, and one had five, all within optimal limits for each technique.

Four descriptions of emergency medical cases were developed for presentation to the panelists. The cases fell within the seven critical care areas of emergency medical services, and were formulated as statements of presenting symptoms. The cases were drawn specifically enough that differential diagnosis should not be a major issue confronting the panelists, and therefore, differences in responses on the cases would not be attributable to differential interpretation of them.

The following are the four case descriptions for which panelists developed process criteria:

Case A: A 48 year old male with a history of angina comes to the emergency room complaining of chest pain more severe than usual, unrelieved by nitroglycerine. The vital signs are:
• BP—140/90, • Pulse—90, • Respiration—16, • Temperature—99.

Case B: A 20 year old male is brought into the emergency room after a severe automobile accident. It is noted that the patient is unconscious and has pupillary inequalities, compound fractures of left humerus and right femur, and a right flank bruise. The vital signs are:
• BP—80/40, • Pulse—110, • Respiration—24, • Temperature—98.6.

Case C: A 19 year old male with a history of episodic wheezing for 10 years relieved by bronchodilators arrives at the emergency room. This is the patient's third emergency department visit for wheezing within the same week. The physical examination reveals that the patient is dusky; has musical wheezing bilaterally, and a prolonged expiratory phase. The vital signs are:
• BP—140/90, • Pulse—120, • Respiration—32, • Temperature—101.

Case D: An 18 year old female is found in her bed unconscious with a number of empty pill bottles by her side. She is brought into the emergency room by her sister and is still unconscious when the doctor sees her. The patient reacts to deep pain stimu-

lation, has normal reflexes, and has pinpoint pupils. The remainder of the physical examination is unremarkable. The vital signs are:

• BP—120/80, • Pulse--100, • Respiration—12, • Temperature—94.

It was felt that additional non-relevant variability in response could be reduced if it were possible to provide all participants with the same general frame of reference concerning Emergency Department capabilities. The following statement therefore accompanied the cases, asking the panelists to assume for purposes of this study that:

> A medical emergency exists when the likelihood of the loss of life, limb or organ is dependent upon the timeliness with which appropriate medical interventions are rendered. Therefore, the orientation of an emergency department is to recognize life threatening or potentially life threatening situations, and to determine the medical care required to eliminate those life threatening situations so as to stabilize and maintain the patient until such time as the patient is transferred to another patient care area or is discharged from the emergency department. (For the purpose of this exercise a holding area within an emergency department should be considered another patient care area).

The panelists were then instructed to list those medical procedures required to stabilize and maintain the patient until such time as the patient is transferred or discharged from the emergency department.

Both the Delphi and Nominal Group Techniques were adapted from their standard forms to be suitable for our purposes here. Nevertheless, the adaptations did retain the essential procedural guidelines of each technique.

In NGT, these decision making steps were followed by each panel for each case:

- silent listing of all process steps
- presentation to the group of one procedure at a time in round robin fashion, by each panelist
- clarification of resultant listings
- first vote by the panel on listed procedures
- discussion of the vote
- second vote
- tabulation of results.

The voting categories in NGT as well as in Delphi were as follows:

- *Essential.* If this procedure were not performed, it would be harmful to the patient.

- *Desirable.* This procedure would be helpful, but the absence of this procedure would not affect the outcome of case management.

- *Not indicated.* Inclusion of this procedure would be neither significant nor detrimental to the management of this case.

- *Contra-indicated.* Inclusion of this procedure would (potentially) be harmful to the patient.

The Delphi Technique consisted of the following steps:

- Round I - Panelists were asked to list all procedures which should be done in order to stabilize and maintain the patient, and to indicate the level of certainty they had that the process measure should be followed;

- Round II - All discrete procedures generated by a given panel for a given case were fed back to all members of that panel, along with the number of times each procedure had been mentioned at the highest level of certainty; each panelist was asked to vote on each process in terms of the above listed voting categories;

- Round III - Panelists were fed back the results of the second round vote for their panel and asked to vote again on each procedure in terms of the established voting categories.

- Tabulation of results.

Finally, because observation of NGT sessions suggested that the group atmospheres in each may have influenced the votes, the reliability of the group votes was tested. Six months after the NGT sessions, each panelist was asked to vote again anonymously on the procedures enumerated by his panel. The result of this vote was compared to the final group vote.

In NGT, all four cases were discussed in a one-day session. Delphi required approximately nine months to complete because of the long response time entailed in the completion of each round and the amount of time required to tally the votes and prepare the ensuing round's forms.

In the implementation of Delphi and NGT, the first NGT vote was equivalent to the Round II Delphi vote, and the second or final NGT group vote was comparable to the Round III Delphi vote. Thus these are the votes which were compared in the analysis of results designed to determine the relative effectiveness of Delphi and NGT.

Findings

Analyses focused on the aggregate results for panels, and only in the rare instance on the individual responses of panelists. Therefore, findings pertaining to how particular physicians with particular characteristics responded to particular cases are not a part of this report. Rather, a comparative analysis is presented of findings which pertain to the relative utility of each decision making technique:

- the qualitative dimensions of the Delphi and NGT processes
- the number of process measures generated using each technique
- the level of consensus achieved with each technique
- the stability over rounds of the consensus reached
- the commonality over panels within each technique of the process measures on which there was consensus, and
- the comparative cost of implementing each of the two techniques.

Qualitative analysis of the Delphi and NGT group processes

An important issue related to the comparative utility of the NGT and Delphi Techniques has to do with what transpired qualitatively in the two methods and how the panelists reacted to the technique in which they were involved. If a technique has deficits with respect to the purpose for which it is intended—namely, the development of consensus—or if it clearly generates resistance to participation among the doctors involved, such difficulties must be taken into account along with the quantitative findings in assessing the results.

Therefore, analysis of the comments of the panelists involved in Delphi and in NGT and review of tapes of the NGT sessions along with observation of two NGT groups was undertaken for the purpose of assessing what went on in the two techniques and how the doctors evaluated each.

The variable which most distinguishes Delphi from NGT qualitatively is the degree to which panelists express satisfaction with the respective techniques. A majority of the comments on the Delphi Technique concerned difficulties with the format of the Delphi questionnaire, specifically with respect to voting on subcategories of criteria. The following are among the comments noted:

> "One should not have to respond to all sub-process measures—choosing one obviates the need to choose another."
>
> "The distinction between essential and desirable does not seem appropriate when choices are various doses, e.g., 0.3 epinephrine or 0.4 epinephrine or 0.3-0.5 epinephrine—*which* is essential? desirable?"

Additional comments from the doctors pertained to factors intrinsic to the Delphi process; for example:

> "I feel insecure about voting on many procedures because many of the decisions in treatment for the listed conditions vary greatly with results of diagnostic studies. . ."
>
> "The process does not provide for interactive progression of therapeutic decision making—therefore results do not reflect true opinions and practices."

This latter remark is representative of numerous comments that some degree of common understanding is lost in Delphi due to the absence of interaction, and suggests by implication that this loss of common understanding has some negative impact on consensus. The inference to be drawn from this position is that NGT which does provide for interaction, would yield a higher degree of common understanding and perhaps therefore a higher level of consensus. However, as is apparent from analysis of findings pertaining to level of consensus, there is no significant difference in the levels of consensus achieved through Delphi and NGT. Nevertheless, the opportunity for collegial interaction provided by NGT but precluded in Delphi, appears to have led to a satisfaction with NGT which Delphi participants did not express.

The somewhat tedious nature of the Delphi Technique may have generated a certain amount of resistance to participating in the exercise, and may thus have been partially responsible for the lengthy response time required for completion of each round. A number of factors therefore point to the necessity of attempting to revise the Delphi format so as to simplify and clarify the requirements of the exercise if future implementation of the technique is contemplated.

In contrast to Delphi panelists, a number of NGT doctors indicated that they were generally pleased with their work on the

cases utilizing this process and enthusiastic about the occasion for interchange that the process provided them. The opportunity to exchange experiences, information and, to some degree, to confirm approaches was similarly considered valuable. It was mentioned that the group approach allowed for a more complete treatment of the development of process measures, because it essentially capitalized on the sum knowledge of the group, while allowing the group to modify their opinions and approaches as they went along. However, there was some feeling expressed that the requirement of adhering to NGT rules created a cumbersome decision making process, and, further, that the inability to follow branching logic in developing process measures was frustrating.

Although all three NGT sessions were conducted by the facilitators in essentially the same manner, and all were implemented within the standard conduct parameters of NGT, differences among the groups emerged with respect to procedures not covered by the agreed upon NGT strategy. The result of such differences was that each group developed characteristic norms. For example, in one panel where there were greater age and status differentials than in other panels, greater deference was shown. This deference, as indicated by the relative lack of verbal disagreement, may have been due to participants' belief in the validity of each other's knowledge and judgment, or it may have been simply polite behavior engendered by the status differences.

Other evidence of this sort suggested that the outcome of NGT was as much a consequence of the group structure and group dynamics as it was of the structured NGT process. Questions therefore arose as to the reliability of the outcome. Given the observation of the composition and dynamics of these groups, given the process measures that were mentioned, as well as those that were completely omitted, and given the great divergence in outcomes between the groups on each case, it seemed possible that the individuals in each of these groups would not define the same procedures as essential *when acting as individuals* as they did when acting as members of a group. Therefore we undertook an additional anonymous vote to test the reliability or stability of the final NGT vote. As is documented later, the final NGT vote proved to be very unstable, with a considerable number of vote changes emerging in the anonymous, reliability round.

Number of procedures generated with each technique

Panelists using the Delphi technique generated, on the average, more than two times as many process measures as did the panelists using NGT. This difference is due largely to differences in the

methods for compiling the listings of process steps. In the NGT round robin and discussion, each panelist is aware of what preceding panelists have suggested, and therefore tends not to present items which are quite similar to previous mentions, or which might be variations of previous items. For example, if "give oxygen" has been presented, the tendency was not to mention the amount of oxygen to be given or the route for administering it as separate items.

In Delphi, on the other hand, because panelists did not interact with each other and therefore had no knowledge in the first round of the suggestions of other panelists, any number of variations on the administration of oxygen, for example, might be proposed. If these mentions were not clearly duplicative of others in that panel, the methodology called for listing them as separate and discrete items, thus leading to longer listings than emerged in NGT.

However, another difference emerged between Delphi and NGT in terms of the numbers of process steps proposed. The pattern of variation is different for the two techniques. For Delphi, there was more variation in the number of steps by case than by panel; but for NGT, there was more variation by panel than by case. The Delphi pattern or variation is not unexpected, given that some cases were clearly more extensive than others. However, the NGT pattern bears comment for it appears to reflect the particular response styles which became normative in each of the panels. For example, some members of one group repeatedly made explicit that their suggestion of a procedure took into account that they were dealing with an "emergency room situation," thus influencing other members not to stray from this precise focus. On all cases, this panel generated the fewest criteria.

It thus seems to be the case that in NGT, group influences affected the decision making process, and consequently, one of the outcomes of the techniques—variations in the numbers of process measures generated by each panel for each case. Although NGT rules are designed to eliminate the effect of group influences, and although NGT rules were closely followed in this exercise, this variable inevitably crept into the decision making process.

Level of consensus achieved

For both NGT and Delphi, consensus was operationally defined as a vote by more than 50 percent of a panel that a given process measure was "essential" to the management of the case under

consideration. "Level of consensus" was defined as the percent of all process measures generated on which each panel achieved consensus for each case. Table 1 presents the findings with respect to levels of consensus reached.

As is clear from the table, the average level of consensus over panels and cases was virtually the same for Delphi and NGT—just under 60 percent. However, there were differences between the techniques for individual case by panel levels; with, for example, Case D having the highest level of consensus in Delphi, and Case A the highest in NGT.

However, again the pattern of variance in levels of consensus achieved by panels on cases was different in Delphi from NGT. One-way analysis of variance on each technique showed the variance to be accounted for at a statistically significant level by the case variable in Delphi [$F(3,8) = 4.24$; $F.05(3,8) = 4.07$]. For NGT, the variance was more adequately though not to a statistically significant degree, accounted for by the panel variable [$F(2,9) = 3.46$; $F.05(2,9) = 4.26$] than by the case variable [$F(3,8) = .89$; $F.05(3,8) = 4.07$].

Two-way analysis of variance to determine whether the decision making method accounted for the variance between techniques showed that neither the method nor case-method interaction explained the variance, but that the case variable did [technique: $F(1,19) = .03$; case: $F(3,19) = 3.47$; $F.05(3,19) = 3.13$]. Thus, variations in level of consensus across techniques are to be accounted for by the case variable and not the decision making method.

Stability of Delphi and NGT votes over rounds

The most notable findings with respect to stability are those which pertain to a comparison of the round to round changes for Delphi and NGT, and the results of the NGT reliability test.

In most instances, the level of consensus increased from the penultimate to the final vote in both Delphi and NGT. However, these changes were more substantial in Delphi than in NGT, where the second group vote was virtually the same as the first. Thus, knowledge of how one's co-panelists voted in the previous round led to change in the Delphi vote but did not have this effect on the NGT group vote.

While there was little change in vote in the NGT group situation from one round to the next, there were dramatic changes in aggregate votes and level of consensus from the final group vote to the anonymous reliability vote. It seems difficult to con-

Table 1
Level of consensus: number and percent of procedures having majority consensus, by panel and case, in final round

		Delphi Round III vote				NGT Round II vote			
	Panel	Number of Procedures suggested	Procedures achieving majority consensus N	%	Average percent achieving majority consensus	Number of Procedures suggested	Procedures achieving majority consensus N	%	Average percent achieving majority consensus
Case A	1	77	47	61.0		56	36	64.3	
	2	76	48	63.2	59.03	18	12	66.7	64.68
	3	70	37	52.9		46	29	63.04	
Case B	1	119	78	65.5		59	37	62.7	
	2	115	67	58.3	60.8	25	19	76.0	62.27
	3	126	75	58.6		52	25	48.1	
Case C	1	75	31	41.3		41	26	63.4	
	2	81	44	54.3	45.53	19	10	52.6	48.57
	3	83	34	41.0		37	11	29.7	
Case D	1	79	54	68.4		38	29	76.3	
	2	74	59	79.7	67.93	29	16	55.2	60.5
	3	61	34	55.7		40	20	50.0	
Average percent achieving majority consensus					58.32				59.01

clude that doctors' *opinions* about how to handle specific cases actually changed to the extent indicated by the data, even given the passage of six months' time. While particular doctors may have different styles of handling cases, the same doctor would tend to handle the same case in a reasonably similar manner from one situation to the next. To the extent that this assumption is true, such a pattern of consistency was not in evidence from the second NGT vote to the reliability vote. Since the changes in consensus levels and actual votes occurred in all panels, there is nothing idiosyncratic about the findings here. Further, since such changes occurred *out* of the group situation, whereas there was minimal change between the two votes taken *in* the group situation, we are led to conclude that, consistent with the data in the qualitative analysis, certain factors led some doctors to vote along lines in the group context which they did not, in many instances, adhere to when they were voting anonymously in the reliability round.

Given that the reliability vote was anonymous and did not feed back to the panelists the final vote in the NGT session, there is little reason to think that it did not reflect the doctors' actual opinions as to how the cases should be managed. It therefore seems reasonable to conclude that group interaction takes place in NGT which produces a result different from the individual response that would otherwise obtain. Further analysis of the reliability vote showed that in the overwhelming majority of instances, individual panelists' vote changes indicated defection from the majority vote of the panel. In some instances, vote changes were very extensive, taking place in the extreme situation on two-thirds of all items listed, with two-thirds of those changes being away from the majority vote of the panel. It seems clear therefore that the NGT group situation tended to constrain the doctors to make judgments which conformed to a group consensus, but to which they chose not to adhere when voting later under conditions of anonymity and no group constraints. If this is the case, it suggests that NGT is a less than reliable technique for reaching a consensus which is actually reflective of each doctor's best judgment as to how emergency cases should be managed.

Commonality of consensus items across panels

It is assumed that there is a certain common denominator of knowledge shared by all emergency medical specialists by virtue of their training and experience, varied though it may be. If this is true, then it can be hypothesized that some of the steps on

which consensus was achieved for each case by each panel using a particular technique will have gained consensus on all panels using that technique. That is, commonality across all panels in each decision making situation will exist among some consensus items for each case. If this hypothesis is supported, then we wished to determine whether there are differences between Delphi panels and NGT panels in the extent to which commonality emerged for each technique.

These questions are critical to this study given that it posits that the greater the commonality evidenced on consensus items among panels of a given technique, the more useful the decision making technique is for facilitating achievement of a meaningful consensus. The basis for this position is the assumption that doctors in emergency medicine share some common knowledge. Since the panels within each technique were not, in terms of significant variables, different from each other, if there were no commonality of consensus items, one would be forced to the conclusion that the particular decision making technique failed to facilitate the specification of consensus items which were regarded as meaningful to most of the emergency medical specialists participating in that technique. The converse of this example therefore indicates that, to the extent that a decision making technique facilitates the development of consensus items which are common to all panels using that technique, and which are therefore regarded as important to case management by a majority of the doctors involved in that technique, then that decision making technique is an effective one. And it is more effective than a decision making technique which does so to a lesser extent.

Table 2 summarizes the results of this analysis. The average percentage of commonality was higher for each case in Delphi than the corresponding percentage was in NGT. Further, a two-way analysis of variance on the data indicated that differences in commonality levels were explained by the method variable exclusively [method: $F(1,19) = 8.99$. $F.01(1,19) = 8.18$. Neither the case variable nor case-method interaction produced statistically significant F scores]. There is thus substantial evidence to conclude that Delphi is the more effective technique for facilitating the development of a consensus which is likely to be shared by other EMS specialists.

It is believed that degree of commonality can serve to indicate the extent to which criteria would be acceptable to other EMS specialists. If the panelists were indeed a representative sample of the universe of outstanding EMS specialists, these results suggest

Table 2
Number and percent of each panel's majority consensus items which are common to all panels, by technique and case

		Panel 1		Panel 2		Panel 3		Average
	Number of consensus items common to all panels	Number of consensus items	Percent common	Number of consensus items	Percent common	Percent of consensus items	Percent common	Percent common
Delphi:								
Case A	16	47	34.0	48	33.3	37	43.2	36.8
Case B	37	78	47.4	67	55.2	75	49.3	50.6
Case C	14	31	45.2	44	31.8	34	41.2	39.4
Case D	20	54	37.0	59	33.9	34	58.8	43.2
Average percent common			40.9		38.55		48.13	42.5
NGT:								
Case A	7	36	19.4	12	58.3	29	24.1	33.9
Case B	9	37	24.3	19	47.3	25	36.0	35.9
Case C	3	26	11.5	10	30.0	11	27.2	22.9
Case D	5	29	17.2	16	31.3	20	25	24.5
Average percent common			18.1		41.7		28.1	29.3

that Delphi has a greater likelihood of being a more useful technique than NGT for developing criteria acceptable to emergency medical specialists generally.

Comparative costs associated with implementing Delphi and NGT

The costs of implementing the two techniques *as employed here* were virtually identical. However, any number of changes in the implementation strategy for either technique would result in different levels of costs for each.

Conclusion

On the basis of the data presented and analyzed, particularly that pertaining to the stability and commonality issues, we conclude that the Delphi Technique is more useful than the Nominal Group Technique for achieving meaningful consensus on the management of emergency medical cases. The data clearly suggest that, due to group influences, NGT generates a consensus which does not reliably reflect the individual medical judgments of the panelists.

We recommend, however, that further research be conducted to determine the comparative acceptability of the criteria developed through each technique. Research along these lines is required if we are to know whether either decision making technique develops consensus on procedures which are judged to be consistently representative of high quality care. If such a determination can be made, that technique can serve as a reliable tool for the development of EMS quality care standards such as are required by PSROs.

Current NCHSR Publications

National Center for Health Services Research publications of interest to the health community are available on request to NCHSR, Publications and Information Branch, 3700 East-West Highway, Room 7-44, Hyattsville, MD 20782 (telephone: 301/436-8970). Mail requests will be facilitated by enclosure of a self-addressed, adhesive-backed mailing label. These publications also are available for sale through the National Technical Information Service (NTIS), Springfield, VA 22161 (telephone: 703/487-4650). PB and HRP numbers in parentheses are NTIS order numbers. Publications which are out of stock in NCHSR are indicated as available only from NTIS. Prices may be obtained from the NTIS order desk on request.

Research Digests

The *Research Digest Series* provides overviews of significant research supported by NCHSR. The series describes either ongoing or completed projects directed toward high priority health services problems. Issues are prepared by the principal investigators performing the research, in collaboration with NCHSR staff. Digests are intended for an interdisciplinary audience of health services planners, administrators, legislators, and others who make decisions on research applications.

(HRA) 76-3144 Evaluation of a Medical Information System in a Community Hospital (PB 264 353, available NTIS only)

(HRA) 76-3145 Computer-Stored Ambulatory Record (COSTAR) (PB 268 342)

(HRA) 77-3160 Program Analysis of Physician Extender Algorithm Projects (PB 264 610, available NTIS only)

(HRA) 77-3161 Changes in the Costs of Treatment of Selected Illnesses, 1951-1964-1971 (HRP 0014598, available NTIS only)

(HRA) 77-3163 Impact of State Certificate-of-Need Laws on Health Care Costs and Utilization (PB 264 352, available NTIS only)

(HRA) 77-3164 An Evaluation of Physician Assistants in Diagnostic Radiology (PB 266 507, available NTIS only)

(HRA) 77-3166 Foreign Medical Graduates: A Comparative Study of State Licensure Policies (PB 265 233, available NTIS only)

(HRA) 77-3171 Analysis of Physician Price and Output Decisions (PB 273 312)

(HRA) 77-3173 Nurse Practitioner and Physician Assistant Training and Deployment (PB 271 000, available NTIS only)

(HRA) 77-3177 Automation of the Problem-Oriented Medical Record (PB 266 881, available NTIS only)
(PHS) 78-3190 Uncertainties of Federal Child Health Policies: Impact in Two States (PB 283 202)
(PHS) 80-3229 Responses of Canadian Physicians to the Introduction of Universal Medical Care Insurance: The First Five Years in Quebec (PB 80-137 979)
(PHS) 79-3231 Israel Study of Socialization for Medicine (PB 293 887)
(PHS) 79-3235 AAMC Longitudinal Study of Medical School Graduates of 1960 (PB 294 689)
(PHS) 79-3238 Some Effects of Quebec Health Insurance (PB 294 097)
(PHS) 79-3261 Medical Education Financing: Issues and Options (PB 80-134 851)
(PHS) 80-3285 A Social and Economic Analysis of Professional Regulation in Optometry

Research Summaries

The *Research Summary Series* provides rapid access to significant results of NCHSR-supported research projects. The series presents executive summaries prepared by the investigators. Specific findings are highlighted in a more concise form than in the final report. The *Research Summary Series* is intended for health services administrators, planners, and other research users who require recent findings relevant to immediate programs in health services.

(HRA) 77-3162 Recent Studies in Health Services Research, Vol. I (July 1974 through December 1976) (PB 266 460)
(HRA) 77-3176 Quality of Medical Care Assessment Using Outcome Measures (PB 272 455)
(HRA) 77-3183 Recent Studies in Health Services Research, Vol. II (CY 1976) (PB 279 198)
(PHS) 78-3187 Criterion Measures of Nursing Care Quality (PB 287 449)
(PHS) 77-3188 Demonstration and Evaluation of a Total Hospital Information System (PB 271 079)
(PHS) 78-3192 Assessing the Quality of Long-Term Care (PB 293 473)
(PHS) 78-3193 Optimal Electrocardiography (PB 281 558)
(PHS) 78-3201 A National Profile of Catastrophic Illness (PB 287 291)
(PHS) 79-3223 Medical Direction in Skilled Nursing Facilities (PB 300 845)
(PHS) 79-3230 Per-Case Reimbursement for Medical Care (PB 294 688)
(PHS) 79-3236 Nurse Practitioners and Physician Assistants: A Research Agenda (PB 294 084)
(PHS) 80-3244 Quality of Ambulatory Care
(PHS) 79-3247 A Demonstration of PROMIS
(PHS) 79-3248 Effects of the 1974-75 Recession on Health Care for the Disadvantaged (PB 80-138 449)
(PHS) 79-3250 Effects and Costs of Day Care and Homemaker Services for the Chronically Ill: A Randomized Experiment (PB 301 171)
(PHS) 80-3265 How a Medical Information System Affects Hospital Costs: The El Camino Hospital Experience
(PHS) 80-3266 Developing a Quality Assurance Strategy for Primary Care (PB 80-162 878)

Research Proceedings

The *Research Proceedings Series* extends the availability of new research announced at key conferences, symposia and seminars sponsored or supported by NCHSR. In addition to papers presented, publications in this series include discussions and responses whenever possible. The series is intended to help meet the information needs of health services providers and others who require direct access to concepts and ideas evolving from the exchange of research results.

(HRA) 77-3154 Advances in Health Survey Research Methods (PB 262 230, available NTIS only)
(HRA) 77-3181 NCHSR Research Conference Report on Consumer Self-Care in Health (PB 273 811)
(HRA) 77-3186 International Conference on Drug and Pharmaceutical Services Reimbursement (PB 271 386)
(PHS) 78-3195 Emergency Medical Services: Research Methodology (PB 279 096)
(PHS) 78-3207 Health Survey Research Methods, Second Biennial Conference (PB 293 492)
(PHS) 78-3208 Drug Coverage Under National Health Insurance (PB 293 468)
(PHS) 79-3209 Health Services Research in Puerto Rico (PB 292 326)
(PHS) 80-3215 Cost Accounting for Pharmaceutical Services (PB 80-157 936)
(PHS) 79-3216 Medical Technology: The Culprit Behind Health Care Costs? (PB 299 408)
(PHS) 79-3225-1 Emergency Medical Services Research Methodology: Workshop 1 (PB 294 048)
(PHS) 79-3225-2 Emergency Medical Services Research Methodology: Workshop 2 (PB 80-142 292)
(PHS) 78-3227 Effects of the Payment Mechanism on the Health Care Delivery System (PB 291 231)
(PHS) 79-3228 A National Conference on Health Policy, Planning, and Financing the Future of Health Care for Blacks in America (PB 292 559)
(PHS) 79-3233 Emergency Medical Services Systems as a Health Services Research Setting (PB 297 102)
(PHS) 79-3254 Medical Technology (PB 80-149 511)
(PHS) 79-3256 Sharing Health Care Costs (PB 80 162 795)
(PHS) 79-3257 Health Facility Reuse, Retrofit, and Reconfiguration (PB 80-142 383)

Research Management

The *Research Management Series* describes programmatic rather than technical aspects of the NCHSR research effort. Information is presented on the NCHSR goals, research objectives, and priorities; in addition, this series contains lists of grants and contracts, and administrative information on funding. Publications in this series are intended to bring basic information on NCHSR and its programs to research planners, administrators, and others who are involved with the allocation of research resources.

(PHS) 79-3220 Emergency Medical Services Systems Research Projects, 1978 (PB 292 558)

(PHS) 80-3271 Emergency Medical Services Systems Research Projects Abstracts, 1979

Research Reports

The *Research Report Series* provides significant research reports in their entirety upon the completion of the project. Research Reports are developed by the principal investigators who conducted the research, and are directed to selected users of health services research as part of a continuing NCHSR effort to expedite the dissemination of new knowledge resulting from its project support.

(HRA) 77-3152 How Lawyers Handle Medical Malpractice Cases (HRP 0014313)

(HRA) 77-3159 An Analysis of the Southern California Arbitration Project, January 1966 through June 1975 (HRP 0012466)

(HRA) 77-3165 Statutory Provisions for Binding Arbitration of Medical Malpractice Cases (PB 264 409, available NTIS only)

(HRA) 77-3184 1960 and 1970 Spanish Heritage Population of the Southwest by County (PB 280 656, available NTIS only)

(HRA) 77-3189 Drug Coverage under National Health Insurance: The Policy Options (PB 272 074)

(PHS) 78-3204 Experiments in Interviewing Techniques: Field Experiments in Health Reporting (PB 276 080, available NTIS only)

(PHS) 79-3210 Telehealth Handbook: A Guide to Telecommunications Technology for Rural Health Care, (PB 292 557, available NTIS only)

(PHS) 78-3211 Emergency Medical Technician Performance Evaluation (PB 285 961)

(PHS) 79-3217-1 Evaluation of Child Abuse and Neglect Demonstration Projects, 1974-77, Vols. 1 and 2 (PB 278 438 and 278 439; vols. 1-12, PB 278 437, the set)

(PHS) 78-3219 Needed Research in the Assessment and Monitoring of the Quality of Medical Care (PB 288 826)

(PHS) 79-3237 A Cost-Effective Approach to Cervical Cancer Detection (PB 295 515)

(PHS) 79-3245 Costs and Benefits of Electronic Fetal Monitoring: A Review of the Literature (PB 294 690)

(PHS) 79-3251 Computer Applications in Health Care (PB 300 838)

(PBS) 79-3258 Effects and Costs of Day Care and Homemaker Services for the Chronically Ill: A Randomized Experiment (PB 80-138 100)

(PHS) 80-3263 Health Status, Medical Care Utilization, and Outcome: An Annotated Bibliography of Empirical Studies; Parts I-IV (PB 284 997, available NTIS only)

NHCES

The *National Health Care Expenditures Study Series* presents information and analyses on critical national health policy issues. Basic data were obtained from the National Medical Care Expenditure Survey, a

statistical picture of how health services are used and paid for. Data Previews give preliminary estimates of key measures.

(PHS) 80-3276 Data Preview 1: Who are the Uninsured?
(PHS) 80-3275 Data Preview 2: Charges and Sources of Payment for Dental Visits with Separate Charges
(PHS) 80-3278 Data Preview 3: Who Initiates Visits to Physicians' Offices?
(PHS) 80-3290 Data Preview 4: Health Insurance Coverage of Veterans
(PHS) 80-3291 Data Preview 5: Charges and Sources of Payment for Visits to Physician Offices
(PHS) 80-3296 Data Preview 6: Waiting Times in Different Medical Settings: Appointment and Office Waits

Policy Research

The *Policy Research Series* describes findings from the research program that have major significance for policy issues of the moment. These papers are prepared by members of the staff of NCHSR or by independent investigators. The series is intended specifically to inform those in the public and private sectors who must consider, design, and implement policies affecting the delivery of health services.

(HRA) 77-3182 Controlling the Cost of Health Care (PB 266 885)

Program Solicitations

(PHS) 81-3292 Grants for Research on Health Promotion and Disease Prevention
(PHS) 81-3293 Extramural Associates Program, 1981

BIBLIOGRAPHIC DATA SHEET	1. Report No. NCHSR 81-3[illegible]	2.	3. Recipient's Accession No.
4. Title and Subtitle COMPARISON OF TWO METHODS FOR THE DEVELOPMENT OF PROCESS MEASURES IN EMERGENCY MEDICAL SERVICES: NCHSR Research Summary Series			5. Report Date March 1981 6.
7. Author(s) Thornell, Carolyn Stein on			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Boone, Young and Associates, Inc. Management Consultants 641 Fifth Avenue New York, New York 1017[illegible]			10. Project/Task/Work Unit No. 11. Contract/Grant No. HSA 23-79-[illegible]
12. Sponsoring Organization Name and Address DHHS, PHS, OASH, National Center for Health Service Research Publication and Information Branch, Room 7-44 3700 East-West Highway Hyattsville, MD [illegible] (tel.: 301/436-[illegible])			13. Type of Report & Period Covered Final Report 1/25/79 - 12/31/[illegible] 14.
15. Supplementary Notes			

16. Abstracts

A comparison was made of the use of two group process techniques, Delphi and Nominal Group Technique, to achieve consensus from panels of Emergency Physicians concerning the steps comprising appropriate care of patients described in brief scenarios. The results suggest that of the two methods, Delphi produces more frequent, greater and more stable consensus, but the process is less satisfying to the physicians involved. The latter attribute was considered a source of delays by physicians in completing the Delphi responses.

17. Key Words and Document Analysis. 17a. Descriptors

NCHSR publication of research findings does not necessarily represent approval or official endorsement by the National Center for Health Services Research or the U.S. Department of Health and Human Services.

Elinor K. Walker, NCHSR Project Officer, 301/436-[illegible].

17b. Identifiers/Open-Ended Terms

Health services research
Emergency Medical Services
Process measures
Delphi
Nominal group technique
Clinical decision making

17c. COSATI Field/Group

18. Availability Statement Releasable to the public. Available from National Technical Information Service, Springfield, VA 2216[illegible]	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 23
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74

U.S GOVERNMENT PRINTING OFFICE 1981 — 341-787:107